VitaSi: A Real-Time Contactless Vital Signs Estimation System

B. Durga Bhavani[1], A. Geethika[2], B. Lakshmi[2], G. Rakshitha[2], Ch. Swathi[2]

[1,2]Department of Information Technology

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

ABSTRACT

In recent decades, measuring vital signs has been a popular topic for many researchers. The system for monitoring vital signs benefits humans in many aspects, such as patient health care and disease prevention for elderly people. As important vital signs, the heart rate and breathing rate, which is also called respiration rate (RR), have been attracting increasing attention and many methods for heart rate (HR) and breath rate (BR) detection have been proposed. The methods for measuring the HR and BR can mainly be divided into contact methods and non-contact methods. Contact methods, as the traditional techniques, are the most widely used measurement methods for signs because of their high accuracy.

Most contact methods rely on Electrocardiography (ECG) or Photoplethysmography (PPG). As the most popular way, an ECG signal containing a considerable amount of useful information on vital signs is obtained by placing adhesive gel electrodes on a person's chest or limbs. PPG is an optical technique that can be used to detect variations in Blood Volume Pulse (BVP) ((i.e., changes in the detected light intensity) in the microvascular bed of tissue. Meanwhile, the changes of the BVP contain useful information on the cardiovascular system, which could estimate the HR and BR. A PPG sensor is usually placed on the fingertip to effectively monitor the variations of the blood flow, because transmitted light can be easily detected. Therefore, PPG has also been a popular contact-based method in recent years. However, these contact methods may cause discomfort and inconvenience for users because users usually need to wear some devices, which requires skin-contact. Therefore, this project proposes a real-time method based on PPG for the contactless estimation of HR and BR.

Keywords: Vital signs, Contact less, HR and BR.

1. INTRODUCTION

In recent decades, measuring vital signs has been a popular topic for many researchers. The system for monitoring vital signs benefits humans in many aspects, such as patient health care and disease prevention for elderly people. As important vital signs, the heart rate and breathing rate, which is also called respiration rate (RR), have been attracting increasing attention. The measurement of vital signs such as heart rate, oxygen saturation (SpO2), respiration rate, body temperature, etc. is an important basic task in biomedical metrology. Conventional devices for such tasks mostly take contact-based measurement approaches. However, contact measurement has several disadvantages. Above all, contact with the body and skin raises the risk of skin irritation and germ contamination [1]. Moreover, contact-based devices significantly limit the freedom of body movement of the patients, and hence, it could lead to severe discomforts. Therefore, the contactless estimation of vital signs using image sensors has continuously gained importance because of its advantages regarding hygiene and patient-friendliness.

For the estimation of heart rate and oxygen saturation, there already exist plenty of works that are based on the photoplethysmography (PPG) measurement of skin using a color camera. In the representative works [2], various approaches have been proposed for heart rate estimation. The core of these approaches is the measurement of the minor temporal variation in skin color. This skin color variation occurs due to different absorption spectra of oxygenated and deoxygenated haemoglobin and should be synchronous with the heartbeat. In these approaches, the first step is the selection of a region of interest (ROI) on the face. The color values of the ROI are measured throughout the frames, and a PPG signal is obtained from them via filtering and signal fusion. Finally, the frequency of the heartbeat correlating to the skin color variation can be derived from the PPG signal [3].

In this context, continuous monitoring of vital signs plays a crucial role in the early detection of conditions that affect the well-being of a patient. The respiratory and heart rates are critical physiological parameters, and by continuously monitoring this information, it is possible to detect drowsiness, sleep apnea, and even depression [4]. However, conventional monitoring devices, usually connected by cables, besides restricting mobility, may cause discomfort and epidermal damage, being therefore inadequate for long-term monitoring. On the other hand, contactless radar-based vital sign monitoring provides several advantages over standard devices. Unlike cameras, radar signals can penetrate through different materials and are not affected by skin pigmentation or ambient light levels [5]. Unlike wearable sensors, radar systems do not require users to wear or carry any additional equipment. In addition, radar devices preserve privacy and can be low power and low cost. These inherent characteristics have drawn the attention of the research community, and a variety of radar types are now being used to address different healthcare applications, including sleep monitoring, life detection and rescue, assisted living, diagnosis, and many others.

2. LITERATURE SURVEY

Hénault et. al [6] presented a practical and fully automatic video-based contactless monitoring framework for the evaluation of patient heart rate is presented. The framework allows the calculation of heart rate in quasi real-time and with a high accuracy, using a consumer grade laptop equipped with a usb-connected webcam. The proposed heart rate assessment algorithm combines state-of-the-art methods in video- and signal-processing for contactless sensing; to attain high accuracy predictions, an additional probabilistic formulation that improves the temporal consistency is proposed and used along with live error rejection. To validate the framework, twenty-two subjects were recorded under varying physiological states (total of 111 two-minute-long videos). Heart rate predictions were compared to medical-grade sensors and showed an accuracy of -0.17 ± 1.81 bpm with a MAE of 0.91 bpm.

Rizal et. al [7] designed and implement the contactless human vital sign parameters measurement including pulse rate (PR) and respiration rate (RR) and also for assessment of human soft biometric parameters i.e., age, gender, skin color type, and body height. Our designed system is based on system on chip (SoC) device which run both FPGA and hard processor while provides real-time operation and small form factor. Experimental results show our device performance has mean absolute error (MAE) 2.85 and 1.46 bpm for PR and RR respectively compared to clinical apparatus. While, for soft biometric parameters measurement we got unsatisfied results on age and gender estimation with accuracy of 58% and 74% respectively. However, for skin color type and body height measurement we reach high accuracy with 98 % and 2.28 cm respectively on both parameters.

Wang et. al [8] proposed a new framework for estimating HRs and BRs by combining a Convolutional Neural Network (CNN) with the Phase-based Video Motion Processing (PVMP) algorithm. The experimental results show that our approach achieves better performance. Meanwhile, we introduce a new challenging dataset with fewer constraints, such as large movements, facial expressions and light interference. In addition, they developed a new Android application, which works in real time and offline, based on a CNN for HR and BR estimations.

Rohmetra et. al [9] prospects of vital signs monitoring for COVID-19 infected as well as quarantined individuals by using DL and image/signal-processing techniques, many of which can be deployed using simple cameras and sensors available on a smartphone or a personal computer, without the need of specialized equipment. We demonstrate the potential of ML-enabled workflows for several vital signs such as heart and respiratory rates, cough, blood pressure, and oxygen saturation. We also discuss the challenges involved in implementing ML-enabled techniques.

Negishi et. al [10] proposed automatic, stable and rapid HR, RR and body temperature measurements using an RGB-thermal sensor and its application for the screening of infectious diseases. This method introduces (1) the sensor fusion approach for the detection of detailed facial landmarks in a thermal image, (2) HR estimation, which introduces tapered window, signal reconstruction and MUSIC and (3) RR estimation, which implements nasal or oral breathing selection using SQI and MUSIC. Moreover, we demonstrated a classification model based on SVM using healthy control subjects and patients with seasonal influenza. The results indicate that the proposed method is indispensable for the high performance of contactless multiple vital sign measurements for infection screening.

Tran et. al [11] aims to build a fully intelligent non-invasive vital-sign signal detection from image analysis in terms of clinical scenarios to extract breathing rate, heart rate, and blood pressure values. The state-of-the-art object detection Yolov3 is used to localize the interesting bounding boxes (chest, face, palm), these ROIs are then tracked by the Mosse algorithm to boost the processing performance. Next, the Pyramidal Lucas-Kanade and remote photoplethysmography techniques are used for extracting the motion signals (breath, pulse) and subtle color change induced by pulse, respectively. Besides, digital signal processing is applied to remove undesired noises for obtaining a clean bio-signal. From experiments conducted, our system can detect breathing rate, heart rate in real-time at a long distance in terms of motion scenarios. Similar to the non-invasive blood pressure estimation system, the proposed deep learning model overcomes the dependence of the high-speed camera in previous works. It satisfies two medical standards (British Hypertension Society and Association for the Advancement of Medical Instrumentations) in estimating Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) with the root mean squared error and mean absolute error for SBP/DBP are 7.942/7.912 mmHg and 6.556/6.372, respectively. The proposed approach estimates blood pressure reliably by only an ordinary webcam with 30 fps in a non-contact continuous manner. Thus, it can be concluded that our system can be applied to healthcare applications.

Yang et. al [12] proposed a contactless vital signs monitoring system, which can measure body temperature (BT), heart rate (HR) and respiration rate (RR) for people with and without face masks using a thermal and an RGB camera. The convolution neural network (CNN) based face detector was applied and three regions of interest (ROIs) were located based on facial landmarks for vital sign estimation. Ten healthy subjects from a variety of ethnic backgrounds with skin colors from pale white to darker brown participated in several different experiments. The absolute error (AE) between the estimated HR using the proposed method and the reference HR from all experiments is 2.70 ± 2.28 beats/min (mean ± std), and the AE between the estimated RR and the reference RR from all experiments is 1.47 ± 1.33 breaths/min (mean ± std) at a distance of 0.6–1.2 m.

Selvaraju et. al [13] provides an overview of data acquisition technology (hardware), image and signal processing (software), accuracy, and application areas. As of today, HR and RR are reliably monitored using RGB cameras in controlled settings only, but other vital signs are still lacking robust and sufficiently precise systems. Subject, camera, setting, and environmental parameters have a significant impact on the accuracy. To overcome these effects, robust algorithms based on advanced signal processing or DL are urgently needed. Additionally, fusion-based approaches bear the potential of enhancing reliability. Different ranges of hardware and software parameter can be investigated to obtain the best possible results for various environments. With respect to the COVID-19 pandemic, a potential application could be to deploy cameras for smart-home-based health monitoring of subjects undergoing quarantine. Recently, the U.S. Food and Drug Administration (FDA) approved smart phone applications for therapy e.g., for tinnitus. In light of this development, smartphone camera-based vital sign monitoring could also be part of digital health.

3. PROPOSED SYSTEM

In the existing system heart and breath rate was calculating based on ECG and other signals but now-a-days due to increasing popularity of deep learning algorithms author has introduced contactless vital signs estimation based on person facial motion. In proposed work we are applying Phase Based Motion Processing to detect motion from faces and then these motions will be extracted as PPG (Photoplethysmography) signals. These signals will get trained with CNN algorithm and then this trained model can be applied on any human face motion signals to predict heart and breath rate.

In proposed work, WEBCAM will read 10 frames which contains human face and then detect face and then extract ROI (region of interest) of face and then extract temporal (current time data) and spatial (current frame data) features and then input this features to Phase-based Video Motion Processing (PVMP) algorithm which will extract PPG signals and then these signals will be input to CNN model to predict heart and breath rate.

PPG Signals

PPG (Photoplethysmography) signals are a type of physiological signal that measures changes in blood volume in peripheral blood vessels. PPG is a non-invasive method used to monitor various cardiovascular parameters, such as heart rate, blood oxygen saturation, and pulse waveforms. The PPG signal is obtained by illuminating the skin with a light source, typically an LED, and detecting the light that is transmitted or reflected back using a photodetector. The light is absorbed differently by oxygenated and deoxygenated blood, leading to variations in the intensity of the detected light. These intensity changes in the PPG signal represent the pulsatile component of blood volume changes associated with each heartbeat.

The PPG signal consists of two main components: the AC (alternating current) component and the DC (direct current) component. The AC component represents the pulsatile changes in blood volume, primarily caused by the cardiac cycle. It reflects the changes in arterial blood volume during systole and diastole. The DC component represents the baseline intensity of the PPG signal and is affected by factors such as tissue absorption and venous blood volume.

PPG signals can be captured from various parts of the body, including the fingertip, earlobe, forehead, or wrist. The choice of measurement location depends on the specific application and the desired physiological parameters to be monitored. PPG signals are widely used in healthcare and wearable devices for monitoring vital signs. They are commonly integrated into fitness trackers, smartwatches, and pulse oximeters. PPG signals provide valuable information about heart rate, blood oxygen saturation levels, and can also be used to detect abnormal cardiac rhythms or variations in arterial stiffness.

Phase-based video motion processing (PVMP)

Phase-based video motion processing is a technique used for analyzing and understanding motion in video sequences. It is based on the idea that motion information can be extracted from the phase component of the video signal. This approach has gained popularity due to its ability to handle various types of motion, including translational, rotational, and complex motions. The process starts by decomposing a video sequence into its constituent frames. Each frame is then transformed into the frequency domain using techniques such as the Fourier transform or the Gabor transform. The phase component of the transformed frames contains information about the local phase shifts in the video signal. By analyzing the phase differences between consecutive frames, phase-based motion processing algorithms can estimate the motion between frames. This estimation is achieved by measuring the local phase shifts and their magnitudes. These measurements can provide valuable information about the direction, speed, and magnitude of the motion.

One of the advantages of phase-based motion processing is its ability to handle complex motion patterns that traditional methods, such as block-based motion estimation, may struggle with. It can effectively capture and represent both global and local motion in a video sequence. Additionally, phase-based techniques are less sensitive to noise and image variations compared to intensity-based methods. Phase-based motion processing finds applications in various areas, including video compression, video surveillance, action recognition, and video editing. It enables tasks such as motion tracking, motion segmentation, and motion-based feature extraction.

CNN Model

According to the facts, training and testing of CNN involves in allowing every source speech via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from [0,1]. Convolution layer is the primary layer to extract the features from a source speech and maintains the relationship between pixels by learning the features of speech by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source speech $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an speech (here $d = 3$, since the source speech is RGB) and a filter or kernel with similar size of input speech and can be denoted as $F\left(k_x, k_y, d\right)$.

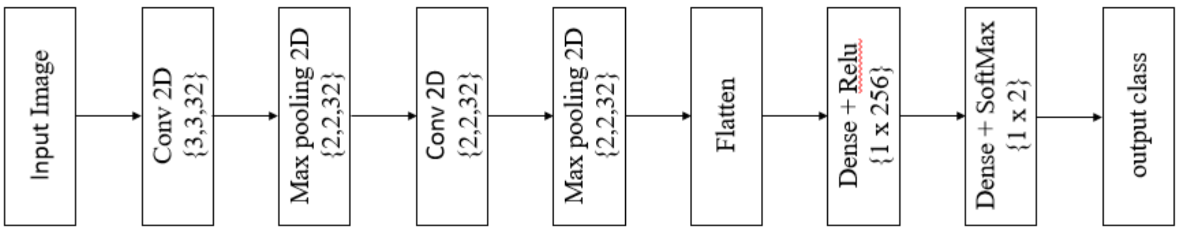


Fig. 1: Representation of convolution layer process

The output obtained from convolution process of input speech and filter has a size of $C\left((x - k_x + 1), \left(y - k_y + 1\right), 1\right)$, which is referred as feature map. An example of convolution procedure is demonstrated in Fig. 2(a). Let us assume an input speech with a size of 5×5 and the filter having the size of 3×3. The feature map of input speech is obtained by multiplying the input speech values with the filter values as given in Fig. 2(b).

ReLU layer: Networks that utilize the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $\mathcal{G}(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input x as follows:

$$\mathcal{G}(x) = \max\{0, x\}$$

Max pooing layer: This layer mitigates the number of parameters when there are larger size speeches. This can be called subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

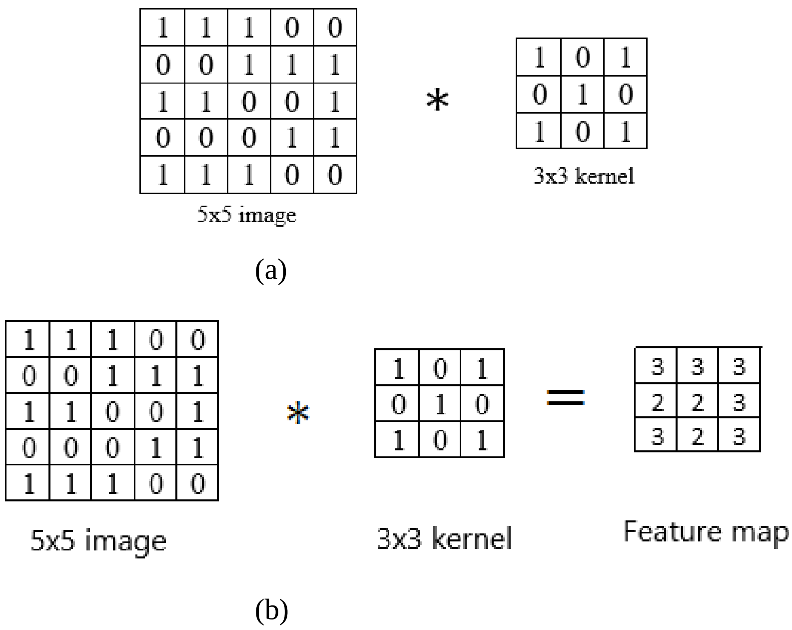


Fig. 2: Example of convolution layer process. (a) a speech with size $\mathbf{5} \times \mathbf{5}$ is convolving with $\mathbf{3} \times \mathbf{3}$ kernel. (b) Convolved feature map.

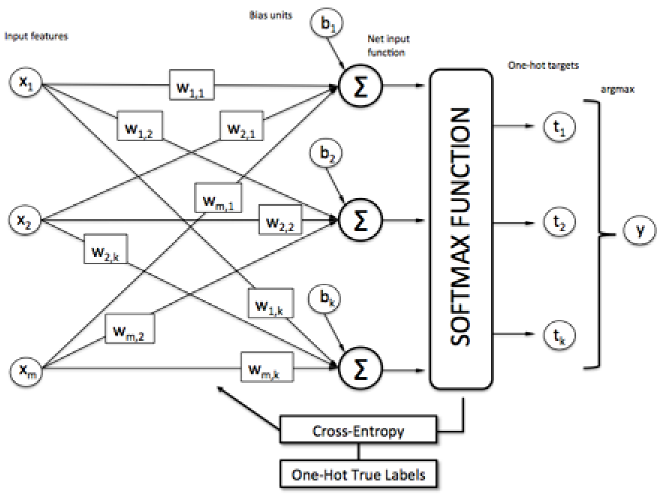


Fig. 3: SoftMax classifier.

SoftMax classifier

Generally, SoftMax function is added at the end of the output since it is the place where the nodes are meet finally and thus, they can be classified. Here, X is the input of all the models and the layers between X and Y are the hidden layers and the data is passed from X to all the layers and Received by Y. Suppose, we have 10 classes, and we predict for which class the given input belongs to. So, for this what we do is allot each class with a particular predicted output. Which means that we have 10 outputs corresponding to 10 different class and predict the class by the highest probability.

In Fig. 4, and we must predict what is the object that is present in the picture. In the normal case, we predict whether the crop is A. But in this case, we must predict what is the object that is present in the picture. This is the place where softmax comes in handy. As the model is already trained on some data. So, as soon as the picture is given, the model processes the pictures, send it to the hidden layers and then finally send to softmax for classifying the picture. The softmax uses a One-Hot encoding Technique to calculate the cross-entropy loss and get the max. One-Hot Encoding is the technique that is used to categorize the data. In the previous example, if softmax predicts that the object is class A then the One-Hot Encoding for:

Class A will be [1 0 0]

Class B will be [0 1 0]

Class C will be [0 0 1]

From the diagram, we see that the predictions are occurred. But generally, we don't know the predictions. But the machine must choose the correct predicted object. So, for machine to identify an object correctly, it uses a function called cross-entropy function. So, we choose more similar value by using the below cross-entropy formula.

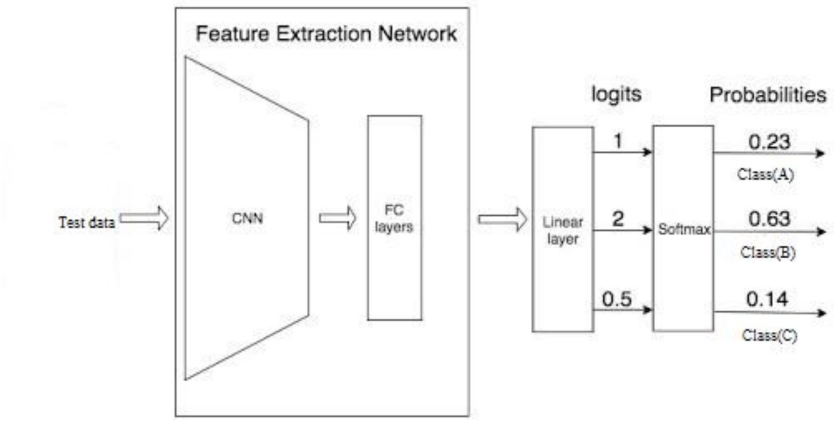


Fig. 4: Example of SoftMax classifier.

In Fig. 5, we see that 0.462 is the loss of the function for class specific classifier. In the same way, we find loss for remaining classifiers. The lowest the loss function, the better the prediction is. The mathematical representation for loss function can be represented as: -

$$LOSS = np.sum(-Y \ * \ np.log(Y\_pred))$$

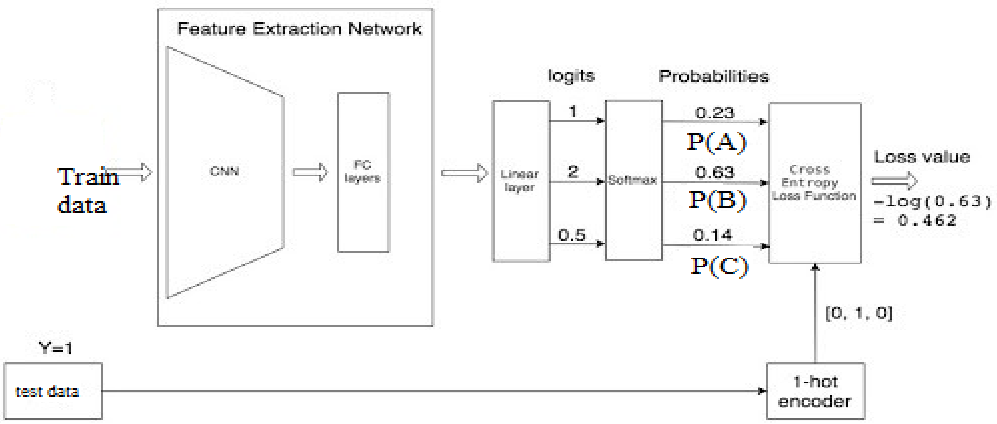


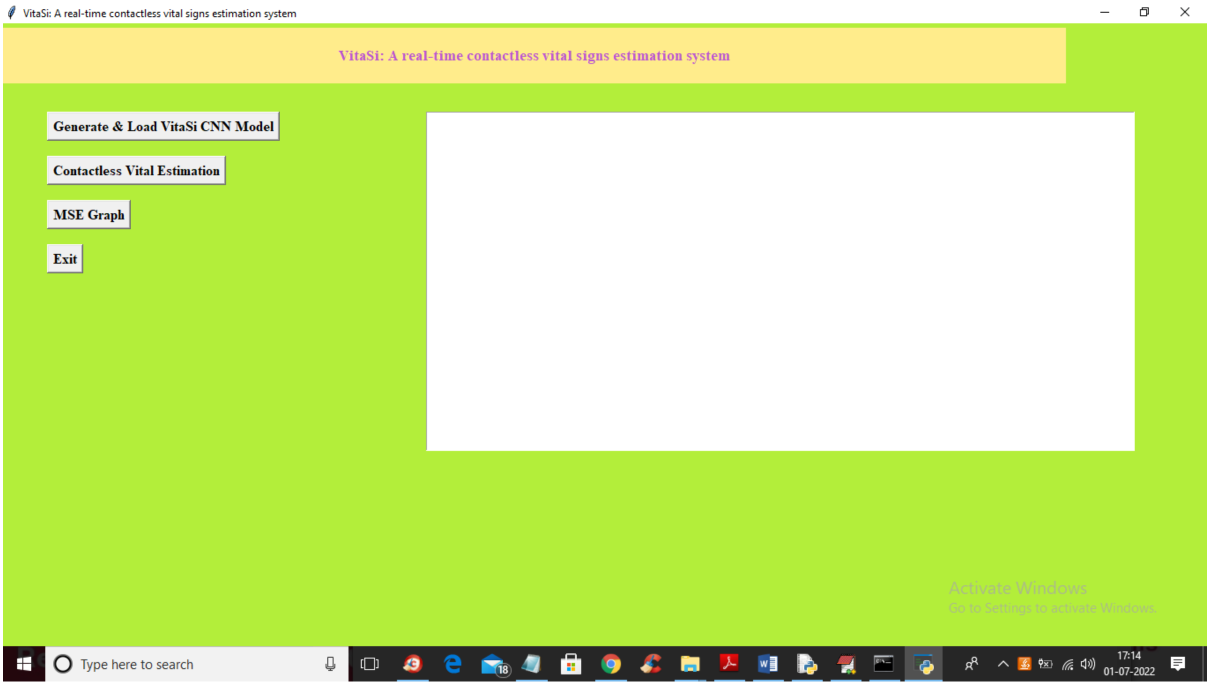
Fig. 5: Example of SoftMax classifier with test data.

4. RESULTS AND DISCUSSION

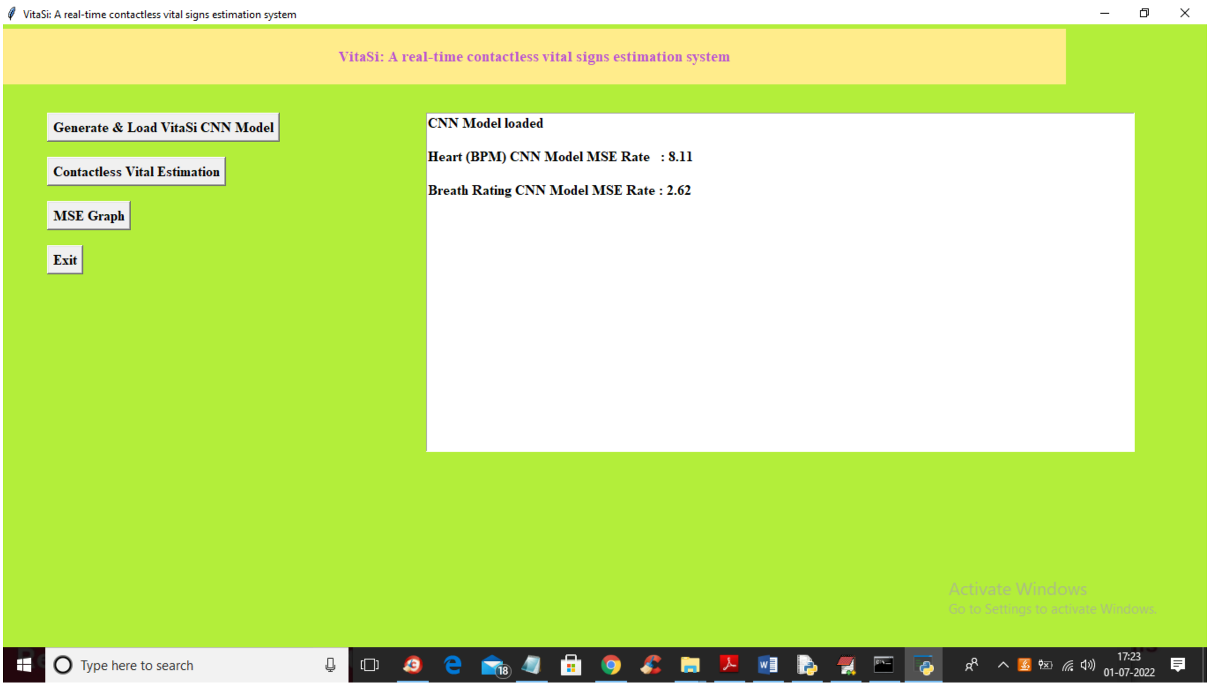
In this project we have used COHFACE dataset which contains 11 person videos and from these videos we extracted faces, heart and breathe rate and input to CNN algorithm to train model.

All these features will be calculated from WEBCAM and to implement this project we have designed following modules.

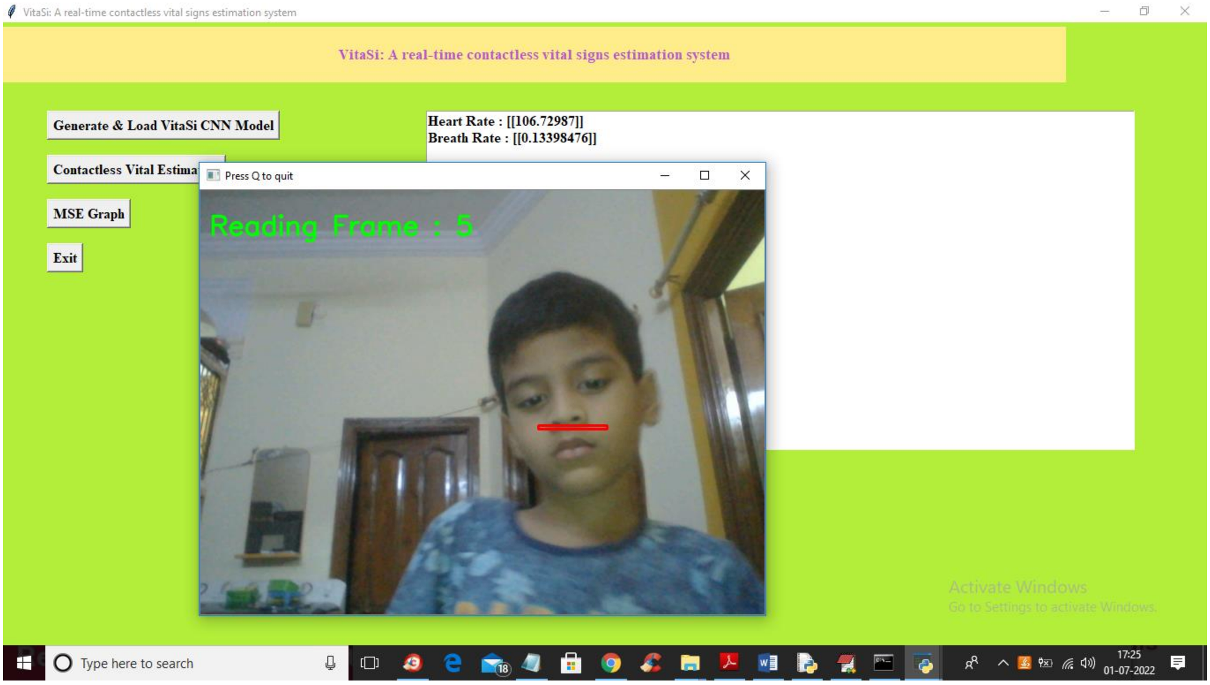
- Generate & Load VitaSi CNN Model: using this module we will load CNN model which will predict heart and breath rate.
- Contactless Vital Estimation: this module will open WEBCAM and then read 10 frames and then estimate heart and breathe rate by extracting PPG signal and employing CNN model.
- MSE Graph: using this module we will display MSE (mean square error) of CNN model for heart and breathe rate prediction. The lower the MSE the better is the prediction model.



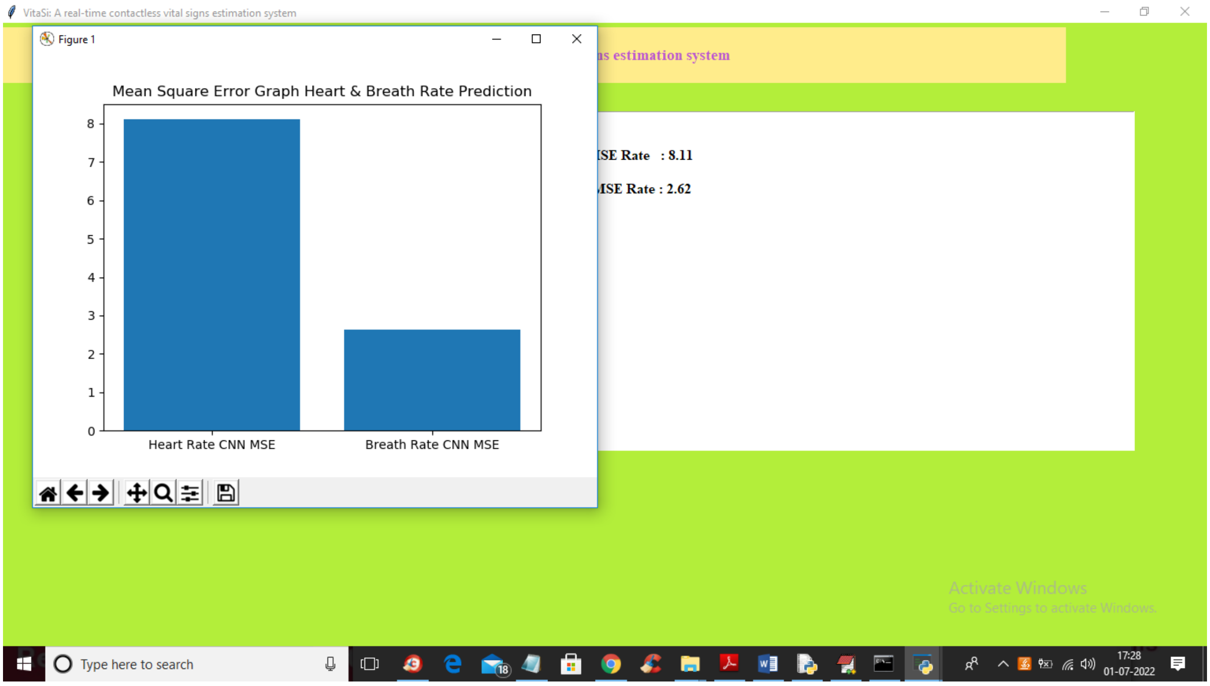
In above screen click on 'Generate & Load VitaSi CNN Model' button to generate and load CNN model and get below screen



In above screen CNN model is loaded and we got Heart Rate MSE as 8.11 and Breathe Rate MSE as 2.62 and now click on ‘Contactless Vital Estimation’ button to start WEBCAM and predict Heart Rate and Breathe Rate



In above screen WEBCAM read 10 frames and then predict heart and breathe rate and this rate will be displayed and updated on TEXT AREA and in above screen we got heart rate as 106 and breathe rate as 0.13 and now click on 'MSE Graph' button to get below graph



In above graph x-axis contains type of prediction and y-axis contains MSE error value of CNN prediction.

Note: if no face shown to WEBCAM, then it will throw exception and stop executing

5. CONCLUSION AND FUTURE SCOPE

This project highlights the importance of measuring vital signs, particularly heart rate (HR) and breathing rate (BR), for patient healthcare and disease prevention. Contact methods, such as Electrocardiography (ECG) and Photoplethysmography (PPG), have been widely used due to their accuracy. However, these methods require physical contact and can cause discomfort for users. To address these limitations, this work proposes a CNN approach for real-time contactless method based on PPG for the estimation of HR and BR. Looking towards the future, there are several potential areas for further exploration. First, validation and accuracy studies can be conducted to compare the contactless PPG-based method with traditional contact methods, ensuring its reliability and effectiveness. Second, refining the algorithms and signal processing techniques can enhance the accuracy of HR and BR estimation using PPG. Third, practical implementation of the contactless method in healthcare and non-medical settings, such as wearable devices or fitness monitoring, can be explored. Lastly, extending the application of the contactless PPG-based method beyond healthcare to areas like stress management and general well-being presents exciting possibilities.

REFERENCES

[1] De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. IEEE Trans. Biomed. Eng. 2013, 60, 2878–2886.

[2] Xu, S.; Sun, L.; Rohde, G.K. Robust efficient estimation of heart rate pulse from video. Biomed. Opt. Express 2014, 5, 1124–1135.

[3] Rapczynski, M.; Werner, P.; Saxen, F.; Al-Hamadi, A. How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2027–2031.

[4] Y. Han, T. Lauteslager, T. S. Lande, and T. G. Constandinou, “UWB radar for non-contact heart rate variability monitoring and mental state classification,” in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2019, pp. 6578–6582.

[5] V. P. Tran, A. A. Al-Jumaily, and S. M. S. Islam, “Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: A comprehensive review,” Big Data Cogn. Comput., vol. 3, no. 1, pp. 1–21, Jan. 2019.

[6] D. Rivest-Hénault, "Quasi Real-time Contactless Physiological Sensing Using Consumer-Grade Cameras," 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), Victoria, BC, Canada, 2021, pp. 193-199, doi: 10.1109/ICHI52183.2021.00038.

[7] A. Rizal, K. -T. Chiang, J. -W. Lin and Y. -H. Lin, "An SoC-Based System for Real-time Contactless Measurement of Human Vital Signs and Soft Biometrics," 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Bandung, Indonesia, 2019, pp. 55-59, doi: 10.23919/EECSI48112.2019.8976953.

[8] Haopeng Wang, Yufan Zhou, Abdulmotaleb El Saddik, VitaSi: A real-time contactless vital signs estimation system, Computers and Electrical Engineering, Volume 95, 2021, 107392, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107392.

[9] Rohmetra, H., Raghunath, N., Narang, P. et al. AI-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges. Computing (2021). https://doi.org/10.1007/s00607-021-00937-7

[10] Negishi, T.; Abe, S.; Matsui, T.; Liu, H.; Kurosawa, M.; Kirimoto, T.; Sun, G. Contactless Vital Signs Measurement System Using RGB-Thermal Image Sensors and Its Clinical Screening Test on Patients with Seasonal Influenza. Sensors 2020, 20, 2171. https://doi.org/10.3390/s20082171

[11] Q. -V. Tran, S. -F. Su, Q. -M. Tran and V. Truong, "Intelligent Non-Invasive Vital Signs Estimation from Image Analysis," 2020 International Conference on System Science and Engineering (ICSSE), Kagawa, Japan, 2020, pp. 1-6, doi: 10.1109/ICSSE50014.2020.9219297.

[12] Yang, F.; He, S.; Sadanand, S.; Yusuf, A.; Bolic, M. Contactless Measurement of Vital Signs Using Thermal and RGB Cameras: A Study of COVID 19-Related Health Monitoring. Sensors 2022, 22, 627. https://doi.org/10.3390/s22020627

[13] Selvaraju, V.; Spicher, N.; Wang, J.; Ganapathy, N.; Warnecke, J.M.; Leonhardt, S.; Swaminathan, R.; Deserno, T.M. Continuous Monitoring of Vital Signs Using Cameras: A Systematic Review. Sensors 2022, 22, 4097. https://doi.org/10.3390/s22114097